

ToxPredict

OpenTox Online Tutorial

Predict the Toxicity of a Compound with
ToxPredict

Apr 20, 2011

IOFA
consultTM

Content

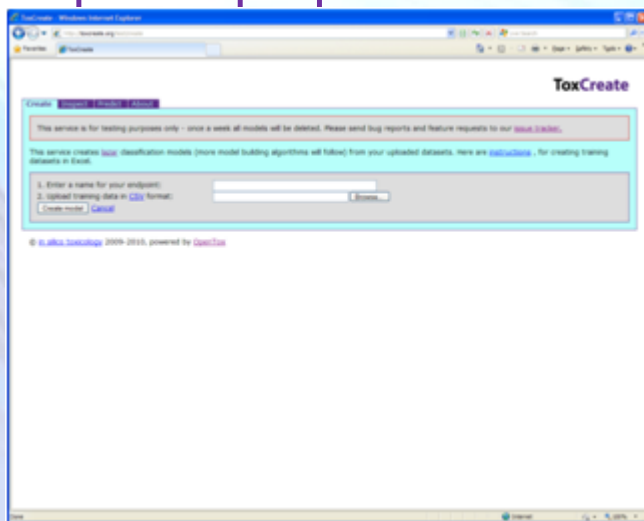
- Overview
- ToxPredict demo
- Exercises 1,2,3
 - Define the structure
 - Select models
 - Obtain predictions
 - Browse toxicity data
- Additional slides (optional)
 - ToxPredict behind the scenes

OpenTox framework

- Distributed Web Services for predictive toxicology
- Several types of Web Services (using REST Web service technology)
- Service types corresponds to following building blocks :
 - Chemical compound;
 - Dataset of chemical compounds and their properties (calculated or measured)
 - Algorithm (descriptor calculation, regression, classification, structural alerts, QC, etc.)
 - Predictive model
 - Report, validation, applicability domain, etc.
- Every object (compound, dataset, algorithm, model, etc.) has an unique web address (e.g. <http://myhost.com/model/bestpredictivemodel>)
- These objects can be created, read, deleted, and updated
- Every object has an RDF (W3C Resource Description Framework) representation, defined in OpenTox ontology (opentox.owl)

Demo applications

- Two end user oriented demo applications, making use of OpenTox webservices, have been developed, deployed and are available for testing - <http://toxcreate.org> and <http://toxpredict.org> ;
- ToxCreate creates models from user supplied datasets;
- ToxPredict uses existing OpenTox models to estimate chemical compound properties



The screenshot shows the ToxPredict web application interface. At the top, it says "ToxPredict". Below that, there is a "Help" button and a "WELCOME, GUEST" message. The main content area displays the chemical structure of 4-methylacetophenone and its properties:

CASRN	122-00-9
EINECS	204-514-8
IUPAC name	1-(4-methylphenyl)ethanone
Chemical Name	4-methylacetophenone
SMILES	CC(=O)c1ccc(C)cc1
Standard InChI	InChI=1S/C9H10O/C=1-3-5-9/6-4-7/2(10)3-8H,1-2H3
Standard InChI key	GNKZMRWRKLTJAYJHFFFAOYSA-N
REACH registration date	30/11/2010

Below the table, there are buttons for "Predictions" and "Datasets", and a "Run All" button. The interface also includes sections for "Physicochemical effects >> Dissociation constant (pKa)", "Environmental fate parameters >> Persistence: Biodegradation", "Ecotoxic effects >> Acute toxicity to fish (lethality)", and "Human health effects".

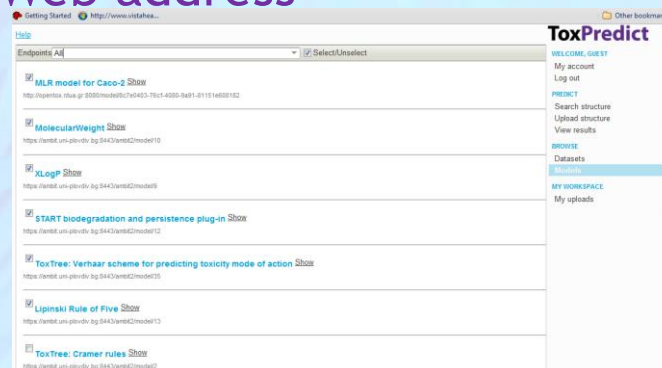
ToxPredict at <http://toxpredict.org>

ToxPredict estimates the chemical hazard of chemical structures. It relies on [OpenTox API-v1.1](#) compliant RESTful webservices. Users can either search the OpenTox prototype database, which includes currently quality labelled data for **~150,000 chemicals**, grouped in number of datasets, or upload their own chemical structure data.

ToxPredict provides access to **16 ready to use models**, addressing **14 different endpoints** (and growing);

ToxPredict uses the following OpenTox webservices: [Compound](#), [Feature](#), [Dataset](#), [Algorithm](#), [Model](#), [Task](#), [Ontology](#), [Validation](#);

Models, available via OpenTox API can be easily integrated in ToxPredict by just publishing its Web address



Exercise 1

- Search for a structure by chemical name
- Run models and obtain predictions
- Explore toxicity data for the selected compound
- Browse a dataset with toxicity data.

Define the structure

Select “Search Structure” .

Enter “Biphenyl” in the query text box. Press “Search” button.

The screenshot displays the ToxPredict web interface. At the top left, there is a [Help](#) link. Below it, a message reads: "Please select the structure(s) for which you would like to apply some OpenTox models." The interface is divided into two main sections: "Draw" and "Search".

The "Draw" section on the left contains a toolbar with various chemical drawing tools (smiles, lines, rings, etc.) and a vertical list of elements: C, N, O, S, F, Cl, Br, I, P, X. Below the toolbar, it says "JME Molecular Editor (c) Peter Ertl" and "JME Editor courtesy of Peter Ertl, Novartis".

The "Search" section on the right has a "Query*" text box containing "biphenyl". Below the text box are radio buttons for "Search mode": "Auto detect" (selected), "Exact structure", "Substructure search", and "Similarity search". A "Search" button is located below these options. At the bottom of the search section, there is a note: "Enter any identifier (CAS, Name, EINECS) or SMILES or InChI or URL of OpenTox compound or dataset. ToxPredict will guess the input type automatically. SMILES may be entered manually into the text field or alternatively use the JME editor to draw the structure."

On the right side of the interface, the "ToxPredict" logo is at the top. Below it, there are navigation links: "WELCOME, GUEST", "My account", "Log out", "PREDICT", "Search structure" (highlighted), "Upload structure", "View results", "BROWSE", "Datasets", "Models", "MY WORKSPACE", and "My uploads".

At the bottom left, there are social media icons for Twitter, LinkedIn, and a "More..." link. At the bottom right, it says "Developed by Ideaconconsult Ltd. 2011".

View the results and run models

The available models are listed under the “Predictions” tab. Click on the “Calculate” link next to each model to obtain predictions.

ToxPredict

WELCOME, GUEST

- My account
- Log out

PREDICT

- Search structure
- Upload structure
- View results**

BROWSE

- Datasets
- Models

MY WORKSPACE

- My uploads

Help





Download
Browse all

CASRN	92-52-4
EINECS	202-163-5
IUPAC name	biphenyl
Chemical Name	biphenyl
SMILES	<chem>c1ccc(cc1)c2ccccc2</chem>
Standard InChI	InChI=1S/C12H10/c1-3-7-11(8-4-1)12-9-5-2-6-10-12/h1-10H
Standard InChI key	ZUOUZKKEUPVFJK-UHFFFAOYSA-N
REACH registration date	30.11.2010

Predictions Datasets

MolecularWeight [Calculate](#)

[Physicochemical effects >> Dissociation constant \(pKa\)](#)

pKa [Calculate](#)

[Physicochemical effects >> Octanol-water partition coefficient \(Kow\)](#)

XLogP [Calculate](#)

[Environmental fate parameters >> Persistence: Biodegradation](#)

START biodegradation and persistence plug-in [Calculate](#)

View the results and run models

The available models are listed under the “Predictions” tab. Click on the “Calculate” link next to each model to obtain predictions. Chemical structure and prediction results can be downloaded via the “Download” link.



ToxPredict

WELCOME, GUEST
My account
Log out

PREDICT
Search structure
Upload structure
View results

BROWSE
Datasets
Models

MY WORKSPACE
My uploads

Help



Download
Browse all

CASRN	92-52-4
EINECS	202-163-5
IUPAC name	biphenyl
Chemical Name	biphenyl
SMILES	<chem>c1ccc(cc1)c2ccccc2</chem>
Standard InChI	InChI=1S/C12H10/c1-3-7-11(8-4-1)12-9-5-2-6-10-12/h1-10H
Standard InChI key	ZUOUZKKEUPVFJK-UHFFFAOYSA-N
REACH registration date	30.11.2010

Predictions Datasets

Run All

MolecularWeight [Calculate](#)

Physicochemical effects >> [Dissociation constant \(pKa\)](#)

pKa [Calculate](#)

Physicochemical effects >> [Octanol-water partition coefficient \(Kow\)](#)

XLogP [Calculate](#)

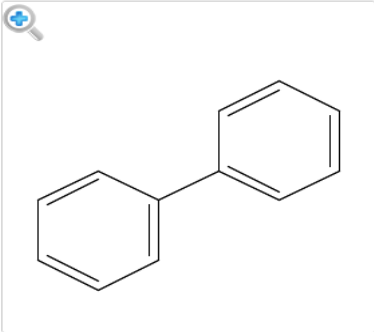
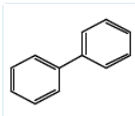
Environmental fate parameters >> [Persistence: Biodegradation](#)

START biodegradation and persistence plug-in [Calculate](#)

Browse toxicity data for the selected compound

The toxicity data for the selected compound could be accessed at the “Datasets” tab. The top level categories shows the endpoint name and the number of datasets (in brackets) containing toxicity information for the selected compounds. Click the “Show” link to view the list of the datasets.

[Help](#)



CASRN	92-52-4
EINECS	202-163-5
IUPAC name	biphenyl
Chemical Name	biphenyl
SMILES	<chem>c1ccc(cc1)c2ccccc2</chem>
Standard InChI	InChI=1S/C12H10/c1-3-7-11(8-4-1)12-9-5-2-6-10-12/h1-10H
Standard InChI key	ZUOUZKKEUPVFJK-UHFFFAOYSA-N
REACH registration date	30.11.2010

Predictions Datasets

AcuteInhalationToxicity (1) [Show](#)

Carcinogenicity (2) [Show](#)

Mutagenicity (4) [Show](#)

[Download](#)
[Browse all](#)

ToxPredict

WELCOME, GUEST

- My account
- Log out

PREDICT

- Search structure
- Upload structure
- View results

BROWSE

- Datasets
- Models

MY WORKSPACE

- My uploads

Browse toxicity data for the selected compound

- Pressing the “Show” link next to the endpoint name list the available datasets.
- The “Show” link next to the dataset title displays the content for the selected compound.
- To browse the entire dataset, use “Browse dataset” link. Note this will bring you to another page, but you could access the current page either via browser “Back” button, or the ToxPredict “View Results” menu on the right.

Predictions Datasets

Show All

AcuteInhalationToxicity (1) [Hide](#)

IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data [Show](#)

Source IRISTR_v1b_544_15Feb2008\IRISTR_v1b_544_15Feb2008.sdf
License: Unknown
See Also http://www.epa.gov/NCCT/dsstox/sdf_iris.html [Browse dataset](#)

Carcinogenicity (2) [Hide](#)

CPDBAS: Carcinogenic Potency Database Summary Tables - All Species [Show](#)

Source CPDBAS_v5d_1547_20Nov2008.sdf
License: Unknown
See Also http://www.epa.gov/NCCT/dsstox/sdf_cpdbas.html [Browse dataset](#)

ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA [Show](#)

Source ISSCAN_v3a_1153_19Sept08.1222179139.sdf
License: Unknown
See Also http://www.epa.gov/NCCT/dsstox/sdf_isscan_external.html [Browse dataset](#)

Mutagenicity (4) [Hide](#)

CPDBAS: Carcinogenic Potency Database Summary Tables - All Species [Show](#)

Source CPDBAS_v5d_1547_20Nov2008.sdf
License: Unknown
See Also http://www.epa.gov/NCCT/dsstox/sdf_cpdbas.html [Browse dataset](#)

Benchmark Data Set for in Silico Prediction of Ames Mutagenicity [Hide](#)

Source tox_benchmark_N6512.sdf
License: Unknown [Browse dataset](#)

Activity	0.00
CAS_NO	92-52-4
DEREK_Example	0.00
DEREK_Pred	0.00
MC_Example	1.00
MC_Pred	0.00
Molecular_Weight	154.21

REFERENCE

JUDSON, PN, COOKE, PA, DOERRER, NG, GREENE, N, HANZLIK, RP, HARDY, C, HARTMANN, A, HINCHLIFE, D, HOLDER, J, MUELLER, L, STEGER-HARTMANN, T, ROTHFUSS, A, SMITH, M, THOMAS, K, VESSEY, JD AND ZEIGER E. TOWARDS THE CREATION OF AN INTERNATIONAL TOXICOLOGY INFORMATION CENTRE. TOXICOLOGY 213(1-2):117-28, 2005

Browse a toxicity dataset

ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA [Hide](#)

Source ISSCAN_v3a_1153_19Sept08.1222179139.sdf

License: Unknown

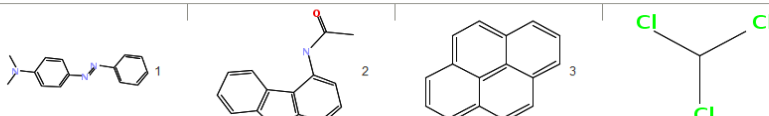
See Also http://www.epa.gov/NCCT/dsstox/sdf_isscan_external.html

[Browse dataset](#)

- Click the “[Browse dataset](#)” link next to the ISSCAN dataset. This will show the “Datasets” page.

Help

Dataset: <https://ambit.uni-plovdiv.bg/8443/ambit2/dataset/9> Page: 0 Show predictions Run models Download



Property

Select Structures	Yes	Yes	Yes	Yes
Dataset				
CAS	60-11-7	28322-02-3	129-00-0	67-66-3
Canc	3.00			3.00
ChemName	4-Dimethylaminoazobenzene	4-Acetylaminofluorene	Pyrene	Chloroform
Formula	C14H15N3	C15H13NO	C16H10	CHCl3
MolWeight	225.30	223.28	202.26	119.38
Mouse_Female_Canc		ND		3.00
Mouse_Female_NTP	ND	ND	ND	ND
Mouse_Male_Canc	3.00	ND		3.00
Mouse_Male_NTP	ND	ND	ND	ND
Rat_Female_Canc	3.00		ND	3.00
Rat_Female_NTP	ND	ND	ND	ND
Rat_Male_Canc	ND	ND	ND	3.00
Rat_Male_NTP	ND	ND	ND	ND
Reference	Toxnet	CPDB	Toxnet	CPDB
SAL	3.00	3.00	3.00	1.00
SMILES	CN(C1=CC=C(C=C1)N=N/C2=CC	O=C(Nc3c2c1cccc1Cc2ccc3)C	c(c(c(cc1)ccc2)c2cc3)(c1ccc4)c34	ClC(Cl)Cl
Substance ID	1.00	2.00	3.00	4.00
Synonyms	Solvent Yellow 2; Butter Yellow; Methyl yellow; DA	4-AAF; 4-acetamidofluorene; N-4-fluorenylacacetamide	Benzo[def]phenanthrene; beta-pyrene; Pyrene[def]ph	Formyl trichloride; methane trichloride; methenyl
TD50_Mouse	ND	ND	NP	90.30

ToxPredict

WELCOME, GUEST

My account
Log out

PREDICT

Search structure
Upload structure
View results

BROWSE

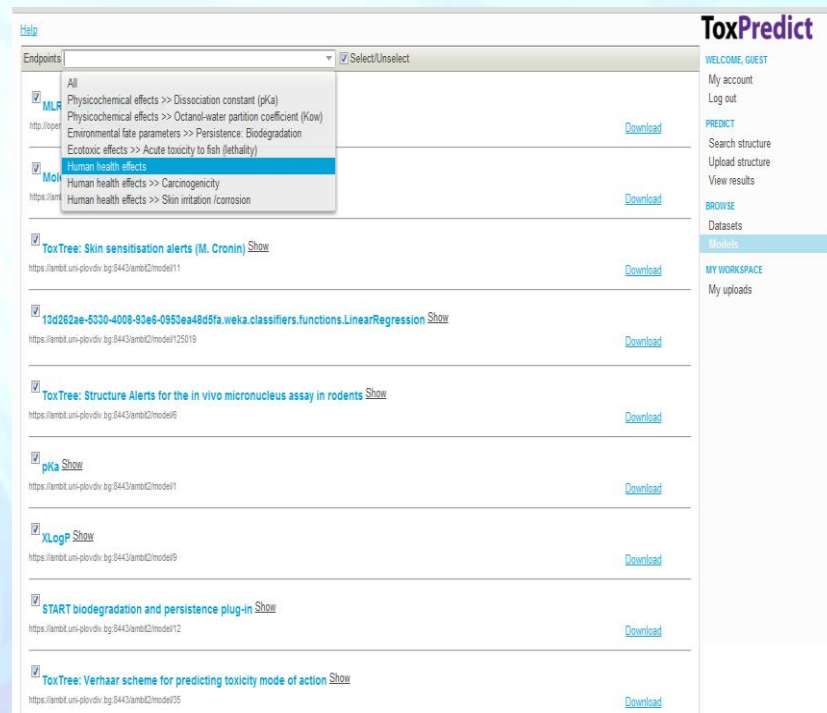
Datasets
Models
MY WORKSPACE
My uploads

Exercise 2

- Select subsets of models to be applied
- Draw a structure
- Search for similar compounds
- View the results and obtain predictions.
- Explore toxicity data for the selected compounds
- Run model predictions in batch mode.
- Browse the results

Select a subset of models

- Select “Models” from ToxPredict menu on the right.
- The list of models appears. Only selected models will be displayed/used when running predictions on other ToxPredict pages.
- Select/Unselect models one by one or by filtered subsets.
 - Use the Endpoints drop down box to filter models by endpoint.
 - The “Select/Unselect” checkbox on the toolbar is applied only on currently visible models.
- Click the “Show” link next to each model to view more information about a model



The screenshot displays the ToxPredict web application interface. On the left, there is a list of models with columns for 'Endpoints', a 'Select/Unselect' checkbox, and a 'Download' link. A dropdown menu is open under the 'Endpoints' column, showing a list of categories such as 'All', 'MLR', 'Mol', and 'ToxTree'. The 'Human health effects' category is currently selected. On the right side of the interface, there is a sidebar menu with options like 'WELCOME, GUEST', 'My account', 'Log out', 'PREDICT', 'Search structure', 'Upload structure', 'View results', 'BROWSE', 'Datasets', 'Models', 'MY WORKSPACE', and 'My uploads'. The 'Models' option is highlighted in the sidebar.

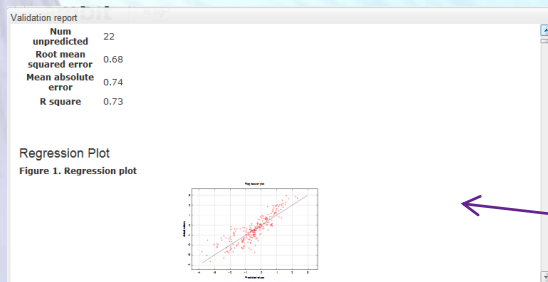
When ready, click on “Search structure” menu on the right.

Model details

Model details includes information about the algorithm, used to derive the model, independent and dependent variables, training dataset, access rights, and links to validation performed and resulted validation reports.

Models are OpenTox web services, derived by learning algorithms, also OpenTox web services.

Validations and validation reports are created by the OpenTox Validation service.



Help

Endpoints: Ecotoxic effects >> Acute toxicity to fish (lethality) Select/Unselect

ECOSAR LC50 fish Hide
http://apps.ideaconsult.net:8080/ambit2/model/238008 [Download](#)

WEKA Linear Regression Model

Created by: N/A

X (Independent variables) Hide
[Column 1](#)

Title: Generated by OpenTox Algorithm [Show](#) Endpoint [Show](#)

XLogP

Y (Observed) [Show](#)

Y (Predicted) [Show](#)

Training dataset [cf47b902-ab36-4901-84a7-67d475a980e1 Show](#)

Source: R738395
License: Unlicense

Algorithm [Show](#)
Algorithm representations

Validation Hide
[Validation](#) · [Validation report](#)

Access rights [Show](#)

ToxTree: Verhaar scheme for predicting toxicity mode of action [Show](#)
http://apps.ideaconsult.net:8080/ambit2/model/35 [Download](#)

ToxPredict

WELCOME, NINA
My account
Log out

PREDICT
Search structure
Upload structure
View results

BROWSE
Datasets
Models

MY WORKSPACE
My uploads

When ready, click on “Search structure” menu on the right.

Define the structure and search mode

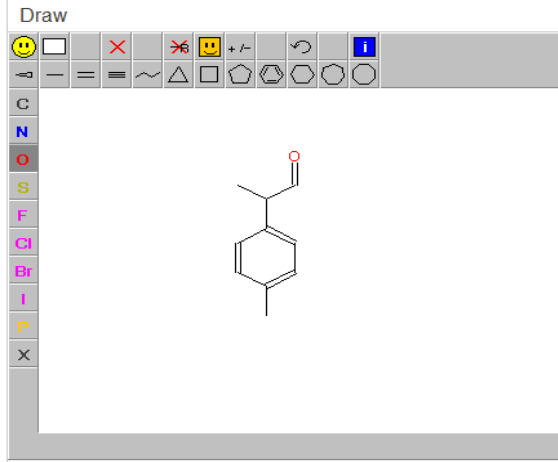
Select “Search Structure” from ToxPredict menu on the right.

- Draw the structure of *2-(4-methyl phenyl) propanal*.
- Select “Similarity search” as a search mode
- Press the “Search” button.

[Help](#)

Please select the structure(s) for which you would like to apply some OpenTox models.

Draw



JME Editor courtesy of Peter Ertl, Novartis

Search

Query *

Search mode

Auto detect

Exact structure

Substructure search

Similarity search

Threshold

Enter SMILES, InChI, or URL of OpenTox dataset or compound. SMILES may be entered manually into the text field or alternatively use the JME editor to draw the structure.

ToxPredict

WELCOME, GUEST

[My account](#)

[Log out](#)

PREDICT

[Search structure](#)

[Upload structure](#)

[View results](#)

BROWSE

[Datasets](#)

[Models](#)

MY WORKSPACE

[My uploads](#)

[Twitter](#) [LinkedIn](#) [More...](#)

Developed by [Ideaconsult Ltd.](#) 2011

View the results and run models

The selected models are listed under the “Predictions” tab. Click on the “Calculate” link next to each model to obtain predictions. Select the “Dataset” tab to verify if there is toxicity data for a compound.

The screenshot displays the ToxPredict web interface. At the top, there is a navigation bar with 'Help' and 'ToxPredict' logos. Below the navigation bar, a row of chemical structures is shown, with the second structure (2-(4-methylphenyl)propanal) highlighted with a blue border. To the right of this row is a sidebar with navigation options: 'WELCOME, GUEST', 'My account', 'Log out', 'PREDICT', 'Search structure', 'Upload structure', 'View results', 'BROWSE', 'Datasets', 'Models', 'MY WORKSPACE', and 'My uploads'. The main content area shows a large chemical structure of 2-(4-methylphenyl)propanal on the left, with a search icon. To its right, a table of identifiers is displayed:

CASRN	99-72-9
EINECS	202-782-0
IUPAC name	2-(4-methylphenyl)propanal
Chemical Name	2-(p-tolyl)propanaldehyde
SMILES	CC(C=O)c1ccc(C)cc1
Standard InChI	InChI=1S/C10H12O/c1-8-3-5-10(6-4-8)9(2)7-11/h3-7,9H,1-2H3/t9-/m1/s1
Standard InChI key	JTZWVKUZBNHSSW-SECBINFHSA-N
REACH registration date	30.11.2010

Below the table, there are two tabs: 'Predictions' and 'Datasets'. A 'Run All' button is located below the tabs. The interface also features several sections for human health effects:

- Human health effects**
Lipinski Rule of Five [Calculate](#)
ToxTree: Cramer rules [Calculate](#)
- Human health effects >> Carcinogenicity**
ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity [Calculate](#)
- Human health effects >> Skin irritation /corrosion**
ToxTree: Skin irritation [Calculate](#)

At the bottom left of the main content area, there are links for 'Download' and 'Browse all'.

As an alternative to inspecting structures one by one, click on the “Browse All” link on the left (just below the selected compound diagram)

View the results and run models

The selected models are listed under the “Predictions” tab. Click on the “Calculate” link next to each model to obtain predictions. Select the “Dataset” tab to verify if there is toxicity data for a compound.

The screenshot displays the ToxPredict web interface. At the top, there is a navigation bar with 'Help' and 'ToxPredict' logos. Below the navigation bar, a row of chemical structures is shown, with the second structure (2-(4-methylphenyl)propanal) highlighted with a blue border. Below this row, a large chemical structure of the selected compound is shown on the left, with 'Download' and 'Browse all' links below it. To the right of the structure, a table of identifiers is provided:

CASRN	99-72-9
EINECS	202-782-0
IUPAC name	2-(4-methylphenyl)propanal
Chemical Name	2-(p-tolyl)propanaldehyde
SMILES	CC(C=O)c1ccc(C)cc1
Standard InChI	InChI=1S/C10H12O/c1-8-3-5-10(6-4-8)9(2)7-11/h3-7,9H,1-2H3/t9-m/s1
Standard InChI key	JTZWVKUZBNHSSW-SECBINFHSA-N
REACH registration date	30.11.2010

Below the table, there are tabs for 'Predictions' and 'Datasets', with a 'Run All' button. Further down, there are sections for 'Human health effects' with links to 'Calculate' for various models: 'Lipinski Rule of Five', 'ToxTree: Cramer rules', 'ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity', and 'ToxTree: Skin irritation /corrosion'.

On the right side of the interface, a sidebar contains the following menu items: 'WELCOME, GUEST', 'My account', 'Log out', 'PREDICT', 'Search structure', 'Upload structure', 'View results', 'BROWSE', 'Datasets', 'Models', 'MY WORKSPACE', and 'My uploads'.

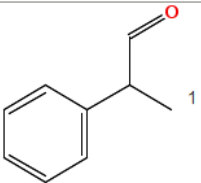
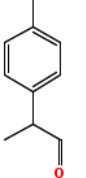
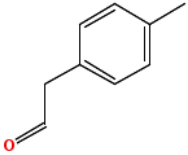
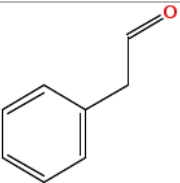
As an alternative to inspecting structures one by one, click on the “Browse All” link on the left (just below the selected compound diagram)

View the results and run models (batch mode)

The similarity search hits are displayed in a table, where columns are compounds, and rows are properties or actions. Page navigation controls are available on the toolbar on the top.

Help

Dataset: <https://ambit.uni-plovdiv.bg:8443/ambit2/query/similarity?sear> Page: 0 Show predictions

Property	1	2	3	4
				
Select Structures	Yes	Yes	Yes	Yes
Dataset				

[Twitter](#) [LinkedIn](#) [More...](#)

Developed by [Ideaconsult Ltd. 2011](#)

ToxPredict

WELCOME, GUEST

- My account
- Log out

PREDICT

- Search structure
- Upload structure
- View results

BROWSE

- Datasets
- Models

MY WORKSPACE

- My uploads

To display model predictions, select “Show predictions” checkbox on the toolbar. This will add as many rows, as necessary to represent the results of applying the previously selected models.

View the results and run models (batch mode)

The similarity search hits are displayed in a table, where columns are compounds, and rows are properties or actions. Page navigation controls are available on the toolbar on the top.

If not all compounds have prediction results, click “[Run models](#)” button to run predictions . The results can be downloaded via the “[Download](#)” link.

The screenshot shows the ToxPredict web interface. At the top, there is a toolbar with buttons for 'Show predictions', 'Run models', and 'Download'. Below the toolbar, four chemical structures are displayed, numbered 1 to 4. Structure 1 is 2-phenylpropanal, structure 2 is 4-methylbenzaldehyde, structure 3 is 4-(4-methylphenyl)butanal, and structure 4 is 3-phenylpropanal. Below the structures is a table of prediction results. The table has columns for each structure and rows for various properties. The 'Human health effects >> Skin irritation /corrosion' section shows that all four structures are predicted to be 'Irritating or Corrosive to skin'. The 'Lipinski Rule of Five' section shows that all four structures have a Lipinski Failure score of 0.00. The 'ToxTree: Cramer rules' section shows that all four structures are predicted to be 'NO' for carcinogenicity and mutagenicity.

Property	1	2	3	4
Human health effects >> Carcinogenicity				
ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity				
For a better assessment a QSAR calculation could be applied	NO	NO	NO	NO
Negative for genotoxic carcinogenicity	NO	NO	NO	NO
Negative for nongenotoxic carcinogenicity	YES	YES	YES	YES
Potential S. typhimurium TA100 mutagen based on QSAR	NO	NO	NO	NO
Potential carcinogen based on QSAR	NO	NO	NO	NO
Structural Alert for genotoxic carcinogenicity	YES	YES	YES	YES
Structural Alert for nongenotoxic carcinogenicity	NO	NO	NO	NO
Unlikely to be a S. typhimurium TA100 mutagen based on QSAR	NO	NO	NO	NO
Unlikely to be a carcinogen based on QSAR	NO	NO	NO	NO
Dataset				
Human health effects >> Skin irritation /corrosion				
ToxTree: Skin irritation				
Skin irritation / skin corrosion	Irritating or Corrosive to skin	Irritating or Corrosive to skin	Irritating or Corrosive to skin	Irritating or Corrosive to skin
Human health effects				
Lipinski Rule of Five				
LipinskiFailures	0.00	0.00	0.00	0.00
LipinskiFailures	0.00	0.00	0.00	0.00
ToxTree: Cramer rules				

Exercise 3

- Upload new chemical structure
- Example file

<http://opentox.org/data/documents/development/tutorialfiles/toxpredict/ToxPredictExercise3/view>

- Run models and obtain predictions
- Explore “My uploads” page

Dealing with “Structure not found”

The “Search structure page” queries the OpenTox dataset service. If the chemical structure is not found in the database, behind the service, it will return “Not found” message.

However, any registered user could upload chemical structures and data via “Upload structure” page.

Help

Please select the structure(s) for which you would like to apply some OpenTox models.

Draw

JME Molecular Editor (c) Peter Ertl
JME Editor courtesy of Peter Ertl, Novartis

Search

Query * C6(C1=CC=CC=C1)=C(C2=CC=CC=C2)C(C3=CC=CC=C3)=C(C4=C

Search mode

- Auto detect
- Exact structure
- Substructure search
- Similarity search

Search

Enter any identifier (CAS, Name, EINECS) or SMILES or InChI or URL of OpenTox compound or dataset. ToxPredict will guess the input type automatically. SMILES may be entered manually into the text field or alternatively use the JME editor to draw the structure.

ToxPredict

WELCOME
Log in

PREDICT
Search structure
Upload structure
View results

BROWSE
Datasets
Models

MY WORKSPACE
My uploads

Not found

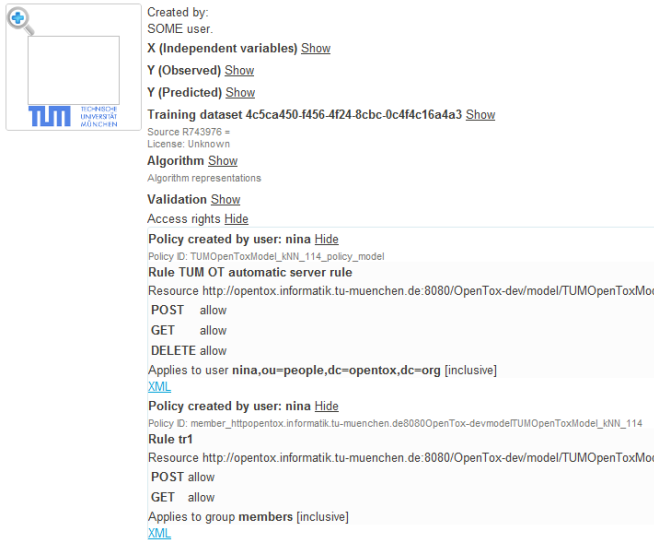
Developed by Ideaconsoft Ltd. 2011

Upload forbidden for anonymous users. Please log in.

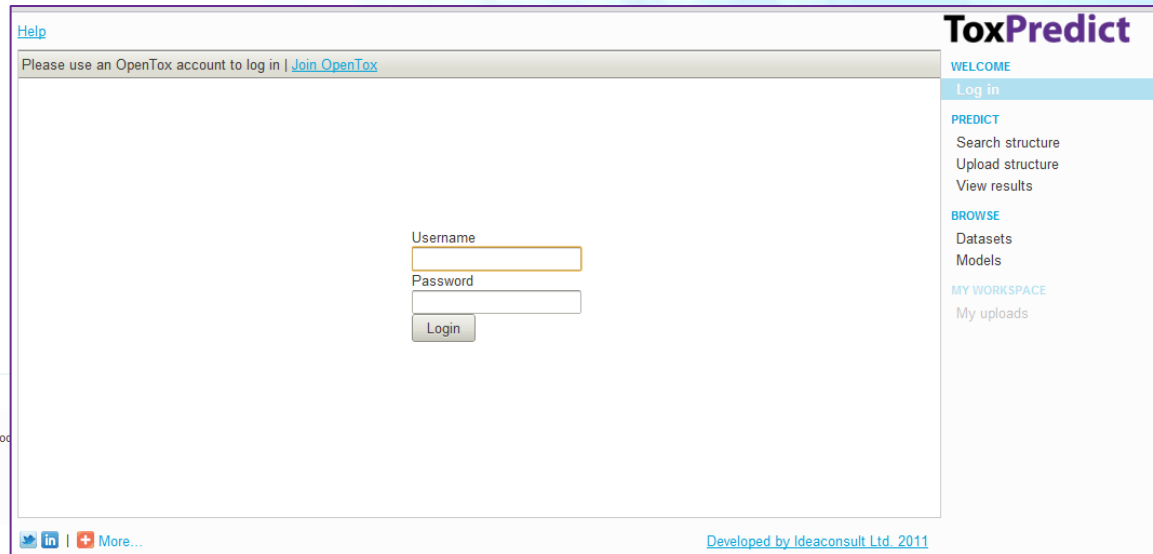
Login in

OpenTox REST web services include both publicly accessible and protected resources. Both data and computational resources could be protected. OpenAM service <http://forgerock.com/openam.html> is used for this purpose. The same users as registered on www.opentox.org site is used across all web services.

OpenTox model created with TUM's kNNregression model learning web service. [Hide](#)
http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/model/TUMOpenToxModel_kNN_114



Created by: SOME user.
X (Independent variables) [Show](#)
Y (Observed) [Show](#)
Y (Predicted) [Show](#)
Training dataset 4c5ca450-4f24-8cbc-0c4f4c16a4a3 [Show](#)
Source R743976 =
License: Unknown
Algorithm [Show](#)
Algorithm representations
Validation [Show](#)
Access rights [Hide](#)
Policy created by user: nina [Hide](#)
Policy ID: TUMOpenToxModel_kNN_114_policy_model
Rule TUM OT automatic server rule
Resource http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/model/TUMOpenToxModel_kNN_114
POST allow
GET allow
DELETE allow
Applies to user nina,ou=people,dc=opentox,dc=org [inclusive]
[XML](#)
Policy created by user: nina [Hide](#)
Policy ID: member_httpopentox.informatik.tu-muenchen.de8080OpenTox-devmodelTUMOpenToxModel_kNN_114
Rule tr1
Resource http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/model/TUMOpenToxModel_kNN_114
POST allow
GET allow
Applies to group members [inclusive]
[XML](#)



Help

Please use an OpenTox account to log in | [Join OpenTox](#)

Username

Password

Login

Developed by Ideacon Ltd. 2011

ToxPredict

WELCOME

Log in

PREDICT

Search structure

Upload structure

View results

BROWSE

Datasets

Models

MY WORKSPACE

My uploads

Upload structures

Switch to “Upload structure” page. The structure, drawn in “Search structure” page is retained, but the options next to it allow to upload the structure drawn or a new file.

Click Upload button when ready.

The screenshot displays the ToxPredict web interface. At the top left, there is a 'Help' link. Below it, a message reads: "Please select the structure(s) for which you would like to apply some OpenTox models." The main area is divided into two panels. The left panel, titled "Draw", contains a chemical structure editor with a toolbar and a vertical element palette (C, N, O, S, F, Cl, Br, I, P, X). A complex polycyclic aromatic hydrocarbon structure is shown in the editor. The right panel, titled "Search", has a "Query*" input field containing the SMILES string: C6(C1=CC=CC=C1)=C(C2=CC=CC=C2)C(C3=CC=CC=C3)=C(C4=CC=CC=C4). Below the input field, there are two radio buttons for "Search mode": "Draw and upload" (selected) and "Upload file with chemical compounds". An "Upload" button is located below the search mode options. The right sidebar features the "ToxPredict" logo and navigation links: "WELCOME, NINA" (My account, Log out), "PREDICT" (Search structure, Upload structure, View results), "BROWSE" (Datasets, Models), and "MY WORKSPACE" (My uploads). At the bottom of the interface, there are social media icons, a "More..." link, and a footer note: "Developed by Ideaconult Ltd. 2011".

View uploaded structures, apply models

Upon upload, the structure(s) will be displayed in the “View results” page.

[Help](#)



[Download](#)
[Browse all](#)

CASRN	992-04-1
EINECS	213-591-7
Chemical Name	3',4',5',6'-tetraphenyl-o-terphenyl
SMILES	<chem>c1ccc(cc1)c2c(c3ccccc3)c(c4ccccc4)c(c5ccccc5)c(c6ccccc6)c2c7ccccc7</chem>
InChI	1S/C42H30
Standard InChI	/c1-7-19-31(20-8-1)37-38(32-21-9-2-10-22-32)40(34-25-13-4-14-26-34)42(36-29-17-6-18-30-36)41(7-1-30H
Standard InChI key	QBHWPVJPWQGYDS-UHFFFAOYSA-N
REACH registration date	31.05.2013

MLR model for Caco-2 [Calculate](#)
MolecularWeight [Calculate](#)
ToxTree: Skin sensitisation alerts (M. Cronin) [Calculate](#)
MLR model for Caco-2 [Calculate](#)
ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents [Calculate](#)

[Physicochemical effects >> Dissociation constant \(pKa\)](#)

ToxPredict

WELCOME, NINA
My account
Log out

PREDICT
Search structure
Upload structure
View results

BROWSE
Datasets
Models

MY WORKSPACE
My uploads

“My uploads” page

Lists the files and structures uploaded by the currently logged-in user. If the dataset service allows protected datasets, the assigned access rights will allow access only to the creator of the resource. These could be changed later , but only by the creator.

The screenshot displays the 'My uploads' page in the ToxPredict interface. At the top, there are tabs for 'My uploads' and 'My models'. The main content area is titled 'Uploaded datasets' and lists two datasets. Each dataset entry includes its URI, creation time, creator's name, and links for 'Predictions' and 'Browse the dataset'. Below the datasets, there is a section for 'Policy created by user: nina' and a 'Rule tr1' with a list of HTTP methods and their permissions. The right sidebar shows the user's name 'NINA', account options, and navigation links for 'PREDICT', 'BROWSE', and 'MY WORKSPACE'. The 'MY WORKSPACE' section is currently expanded to show 'My uploads'.

The dataset service, used in the tutorial , doesn't allow protected resources. All structures uploaded today will become publicly accessible!

“My uploads” page

The uploaded resources will be retained after log on. They could be reused on subsequent logon.

- Click on “Predictions” link to open the datasets into “View Results” page or
- Click on “Browse dataset” to view the structures and data in table mode.

Help

My uploads My models

Uploaded datasets [Hide](#)

Datasets uploaded via ToxPredict

Drawn and uploaded via ToxPredict

Dataset URI: <https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/74709>

Bookmark created on :4/16/11 9:37 PM
Bookmark created by nina

[Predictions](#) [Browse the dataset](#)

Dataset columns [Show](#)

Dataset access rights [Show](#)

Drawn and uploaded via ToxPredict

Dataset URI: <https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/74711>

Bookmark created on :4/20/11 1:05 PM
Bookmark created by nina

[Predictions](#) [Browse the dataset](#)

Dataset columns [Show](#)

Dataset access rights [Hide](#)

Policy created by user: [nina](#) [Hide](#)

Policy ID: <https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/74711> GETPUTPOSTDELETE

Rule tr1

Resource <https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/74711>

POST allow
GET allow
DELETE allow
PUT allow

Applies to user [nina](#) [inclusive]

[XML](#)

ToxPredict

WELCOME, NINA

My account
Log out

PREDICT

Search structure
Upload structure
View results

BROWSE

Datasets
Models

MY WORKSPACE

My uploads

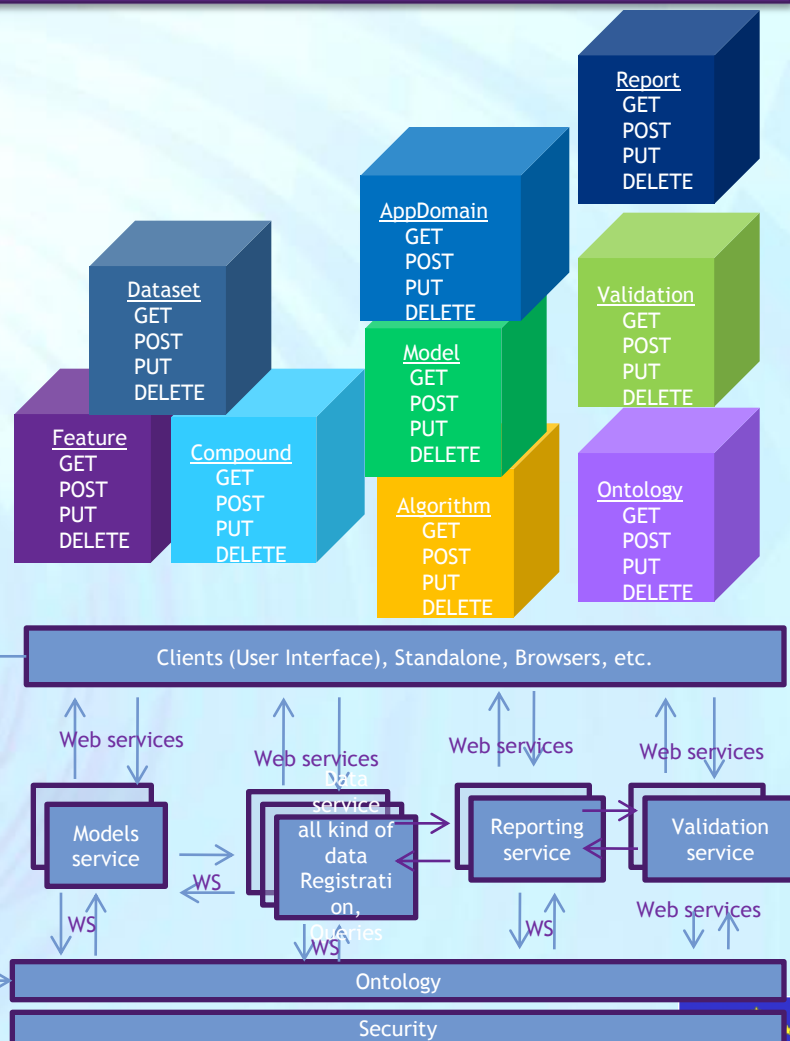
End of exercises.

Behind the scenes

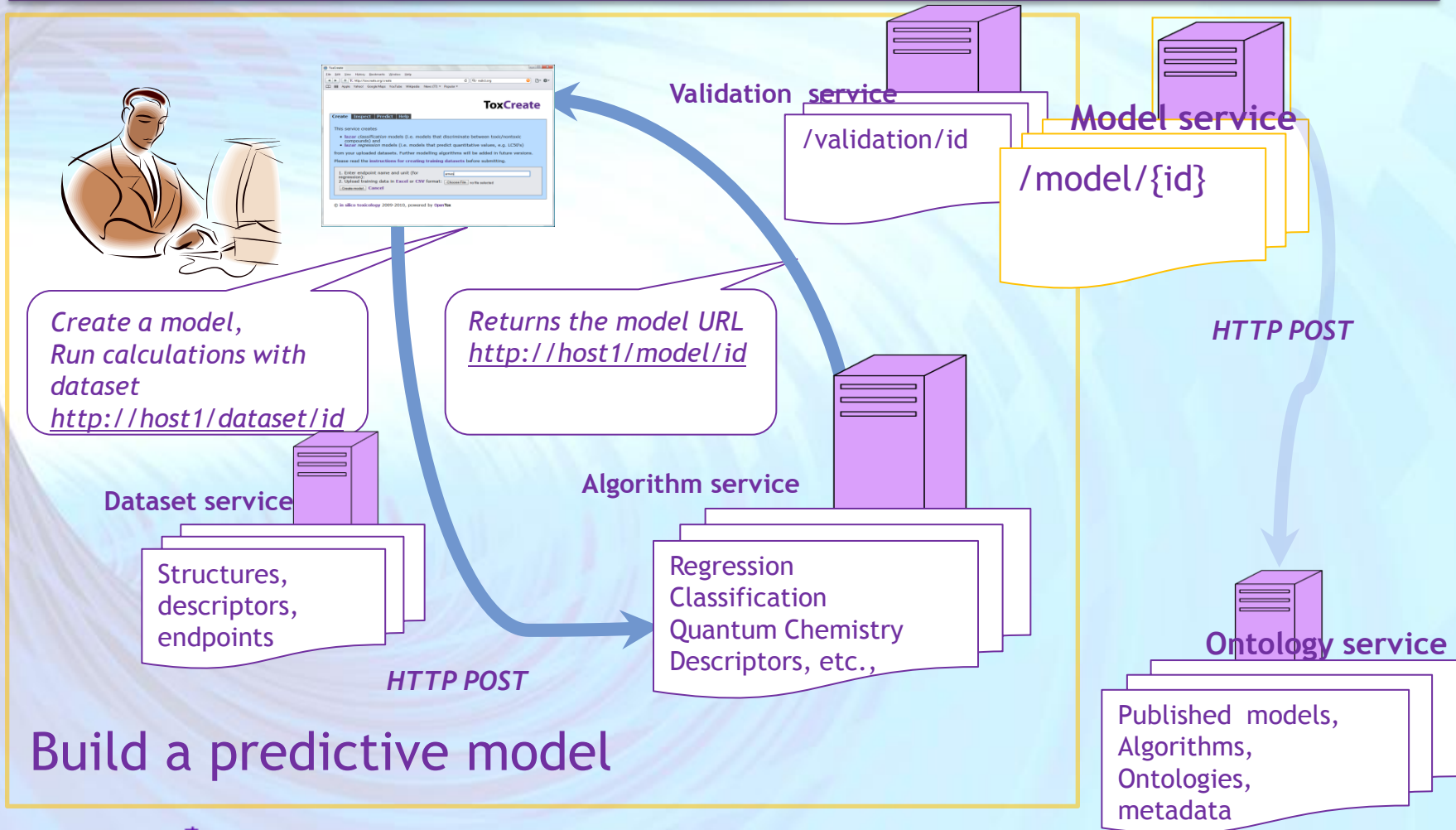
The following slides present a more in-depth overview of ToxPredict's workflow and provide some more details on its interactions with OpenTox webservices which are taking place behind the scenes and without requiring any end-user intervention;

OpenTox framework

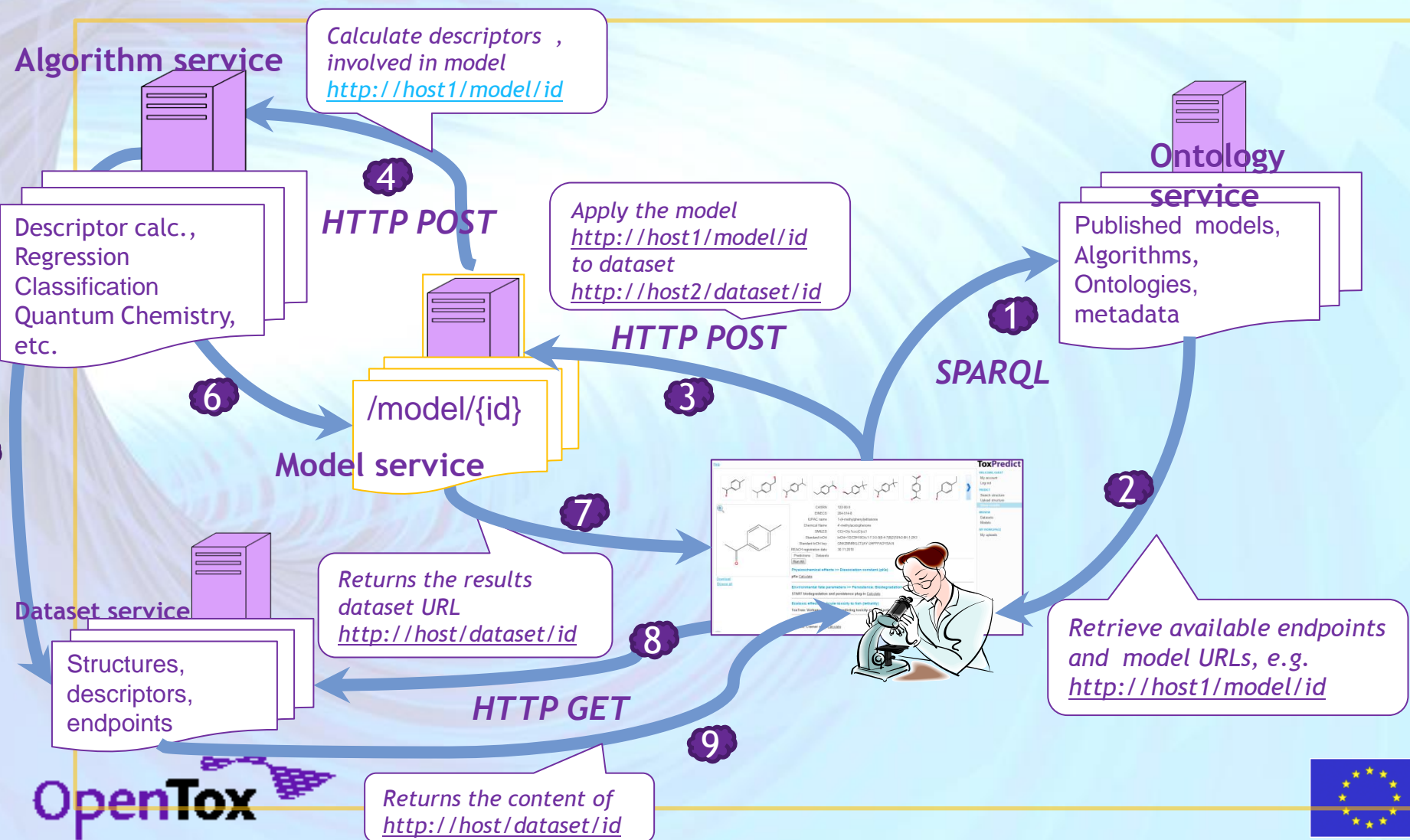
- OpenTox API
 - The way applications talk to each other
 - The way developers talk to applications
 - <http://opentox.org/dev/apis/api-1.1>
- The basic building blocks:
 - data, chemical structures, algorithms and models.
- Functionality offered
 - build models,
 - apply models,
 - validate models,
 - access and query data in various ways.
- Technologies
 - REST style web services
 - RDF for description of resources
 - Links to existing and newly developed ontologies (mainly to describe metadata) about resources



Build a predictive model (ToxCreate)



Apply predictive models (ToxPredict)



ToxPredict: Step 1 (Select structure(s))



Find structure by name, registry number, SMILES, InChI, structure, substructure, similarity...

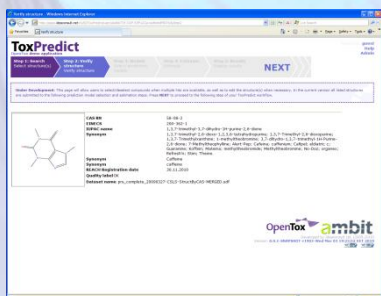


ToxPredict
Web
Application

OT Dataset API *HTTP GET*

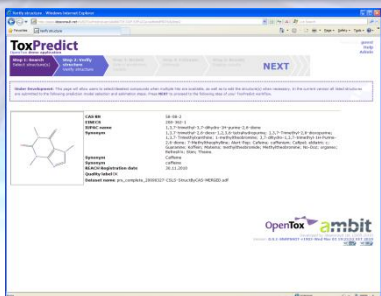
OT Dataset
Service

text/uri-list,
application/rdf+xml,
chemical/x-daylight-smiles
chemical/x-mdl-sdf,...



Here is the list of structures as
URI links, RDF, MOL or SMILES.

ToxPredict: Step 2 (Verify structure(s))



Select and/or edit structure(s)

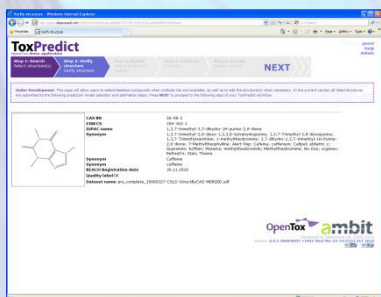


ToxPredict
Web
Application

OT Dataset API *HTTP GET*
(*HTTP POST* for structure editing)

OT Dataset
Service

text/uri-list,
application/rdf+xml,
chemical/x-daylight-smiles
chemical/x-mdl-sdf,
image/png



Here is the list of structures as
URI links, RDF, MOL, SMILES or
images.

ToxPredict: Step 3 (Select model(s))



What prediction models are available? Is there a model for endpoint X?

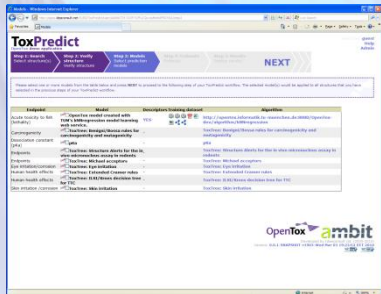
HTTP GET SPARQL query

OT Ontology Service

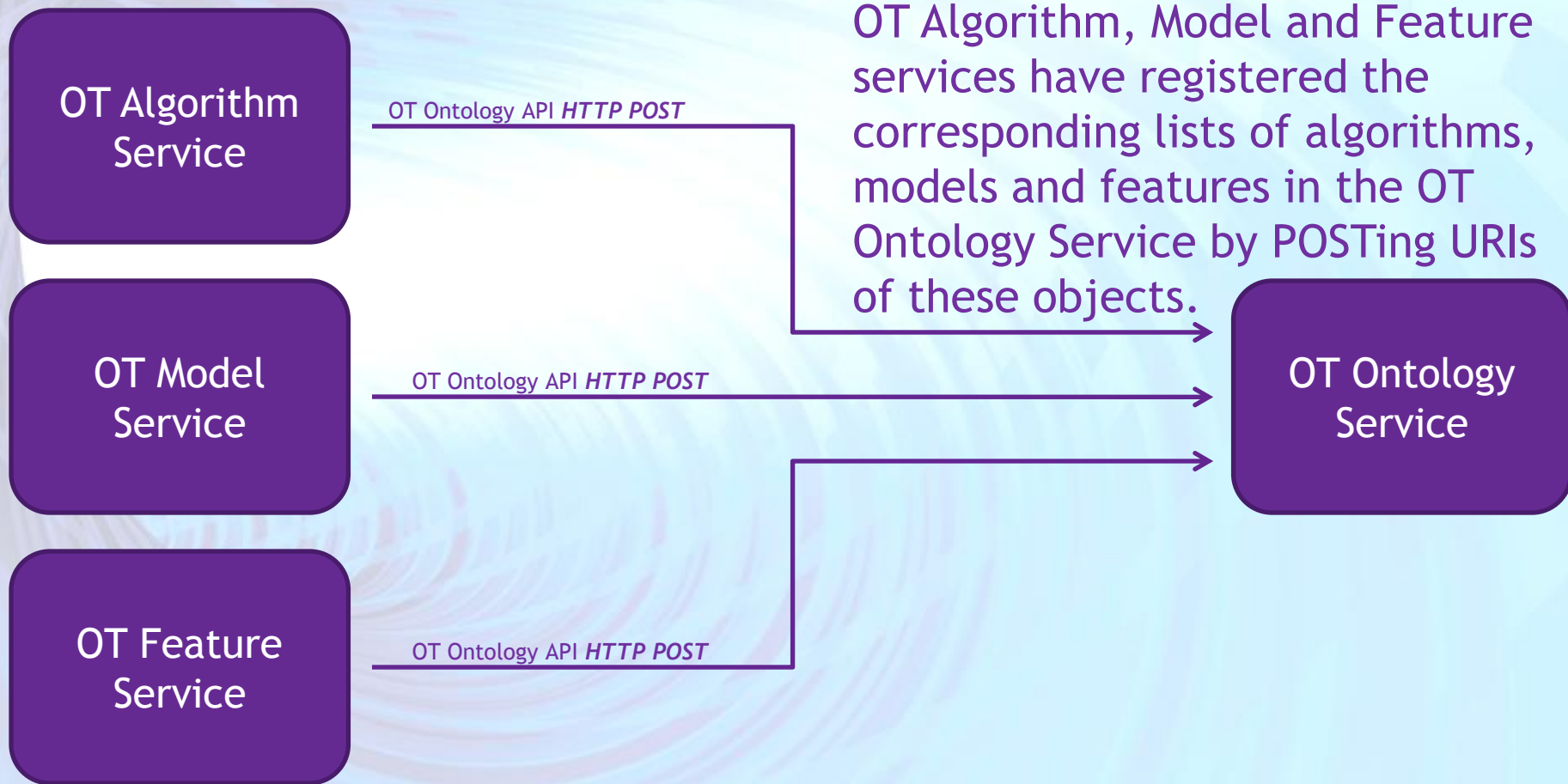
ToxPredict Web Application

application/sparql-results+xml

Here is the list of model URIs and related endpoints and algorithms in SPARQL format.

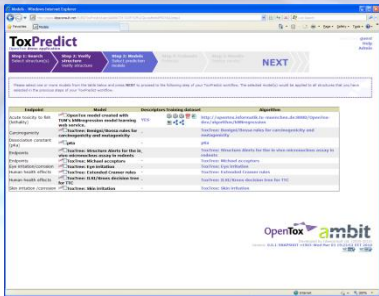


ToxPredict: Step 3 (behind the scenes)



OT Algorithm, Model and Feature services have registered the corresponding lists of algorithms, models and features in the OT Ontology Service by POSTing URIs of these objects.

ToxPredict: Step 4 (Estimate)



Run the selected models.

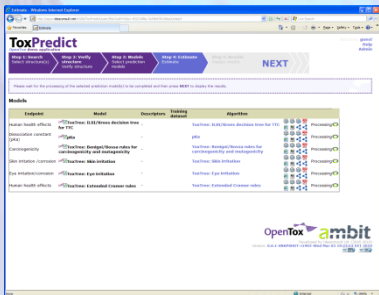


ToxPredict Web Application

OT Model API *HTTP POST* with parameter dataset URI from steps 1-2

OT Model Service

HTTP code 202 "Accepted"
Model task URI in HTTP:location header

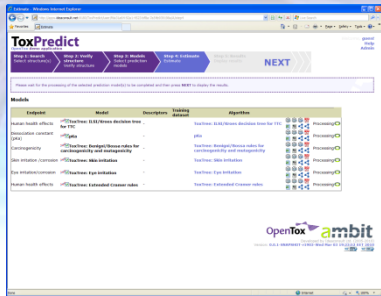


The calculation will take a while; here is a task URI, which can be queried for processing status.

Create a new task.

OT Task Service

ToxPredict: Step 4 (Estimate)



ToxPredict Web Application

Is the task completed?

OT Task API *HTTP GET* on task URI

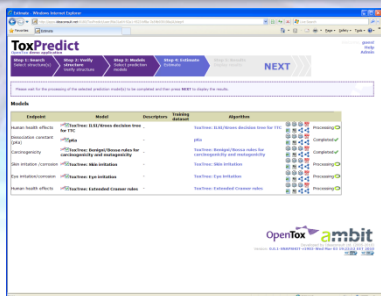
OT Task Service

HTTP code 303 "Redirect"
Dataset task URI in
HTTP:location header



Not yet, but processing has finished and the results have been posted to the OT Dataset service; here is a task URI for the dataset import.

ToxPredict: Step 4 (Estimate)



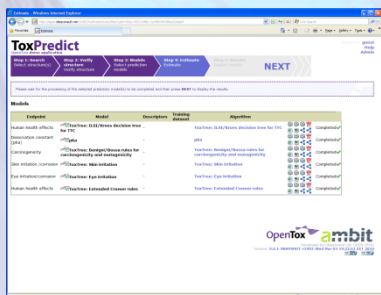
ToxPredict Web Application

Is the task completed?

OT Task API *HTTP GET* on task URI

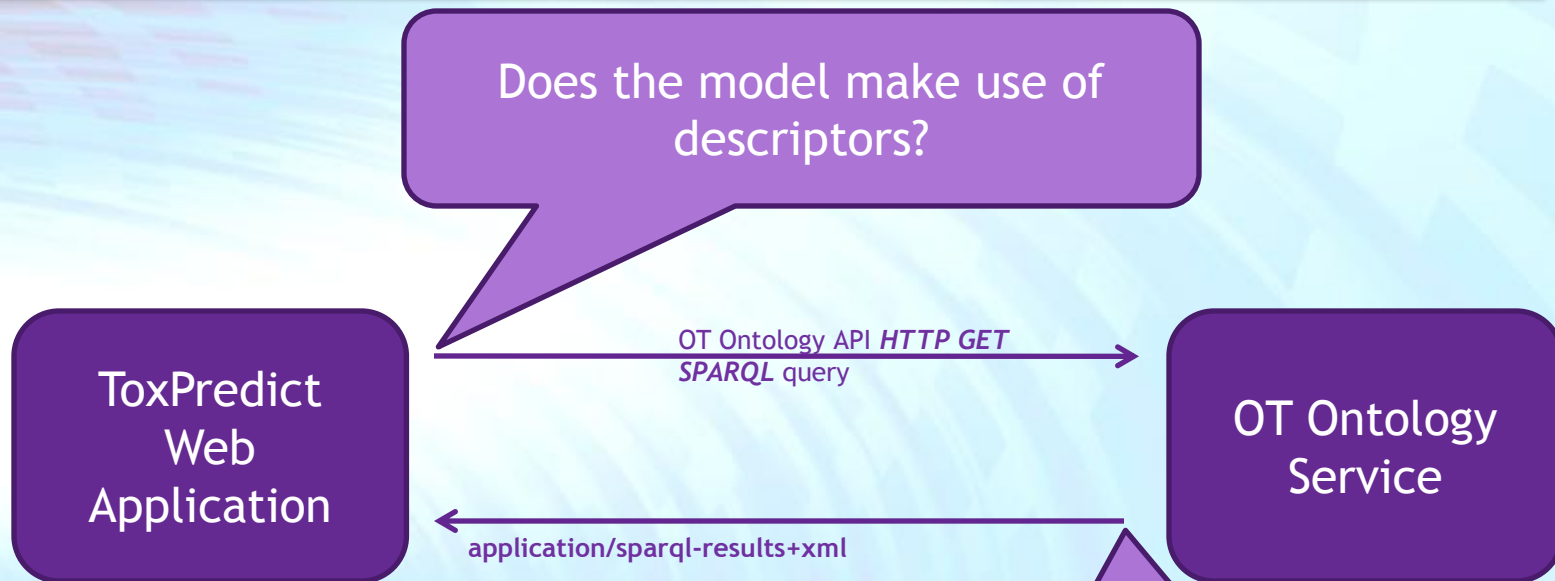
OT Task Service

HTTP code 200 "OK"
Dataset URI in
HTTP:location header



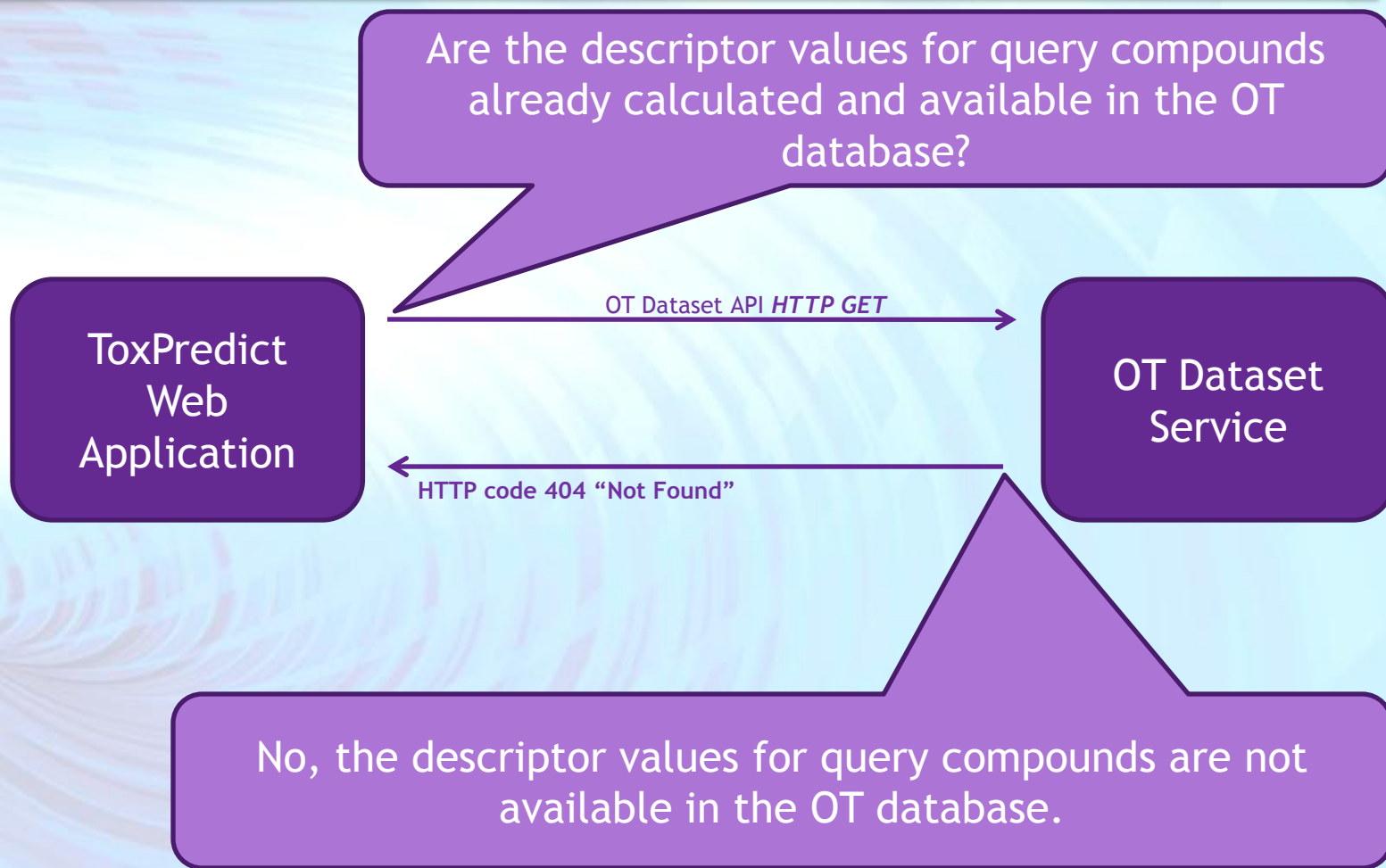
Yes, here is the dataset URI of the results.

ToxPredict: Step 4 (behind the scenes)

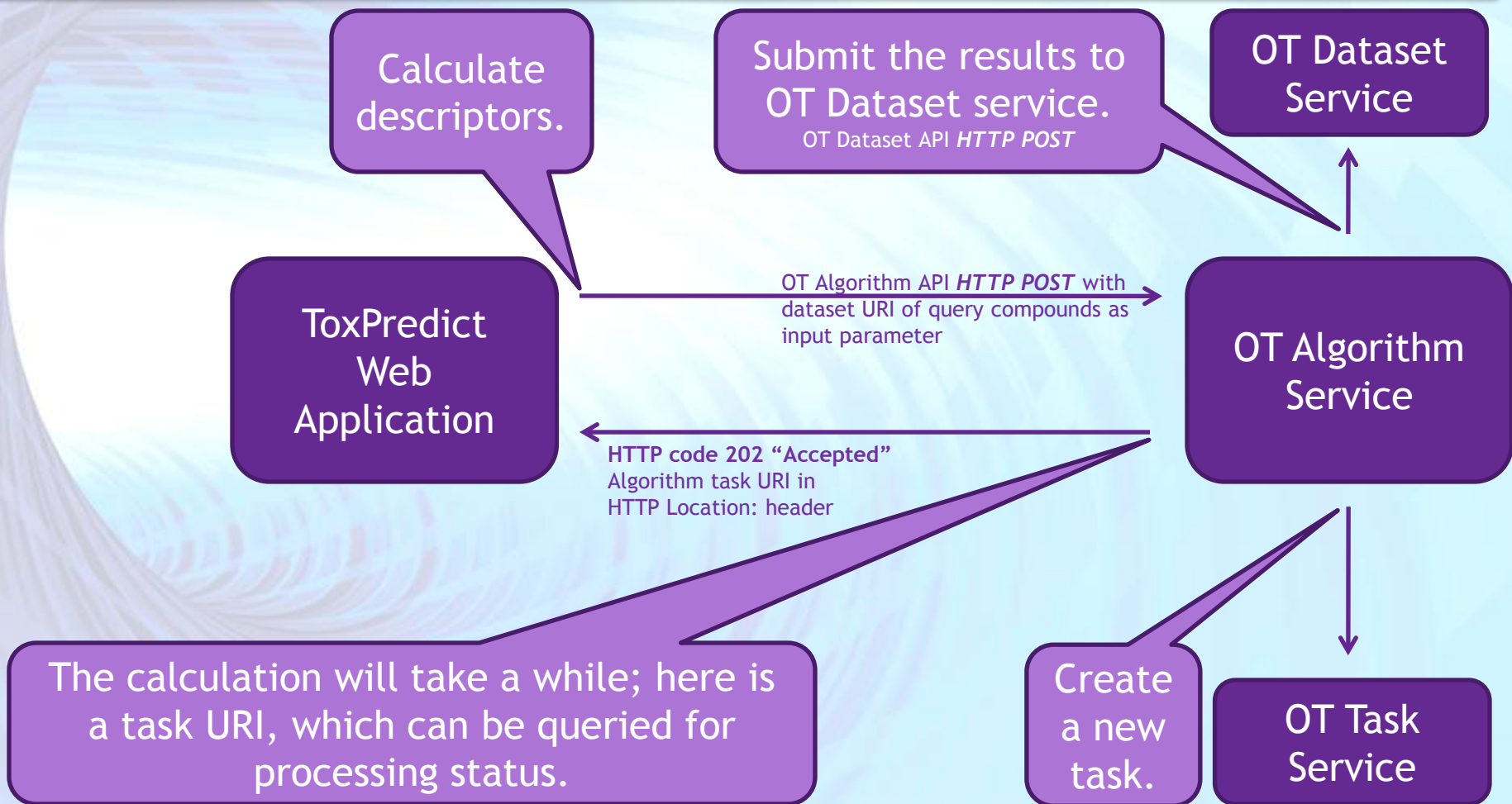


Yes, here is a list of features and algorithms, used to calculate descriptors for the training set.

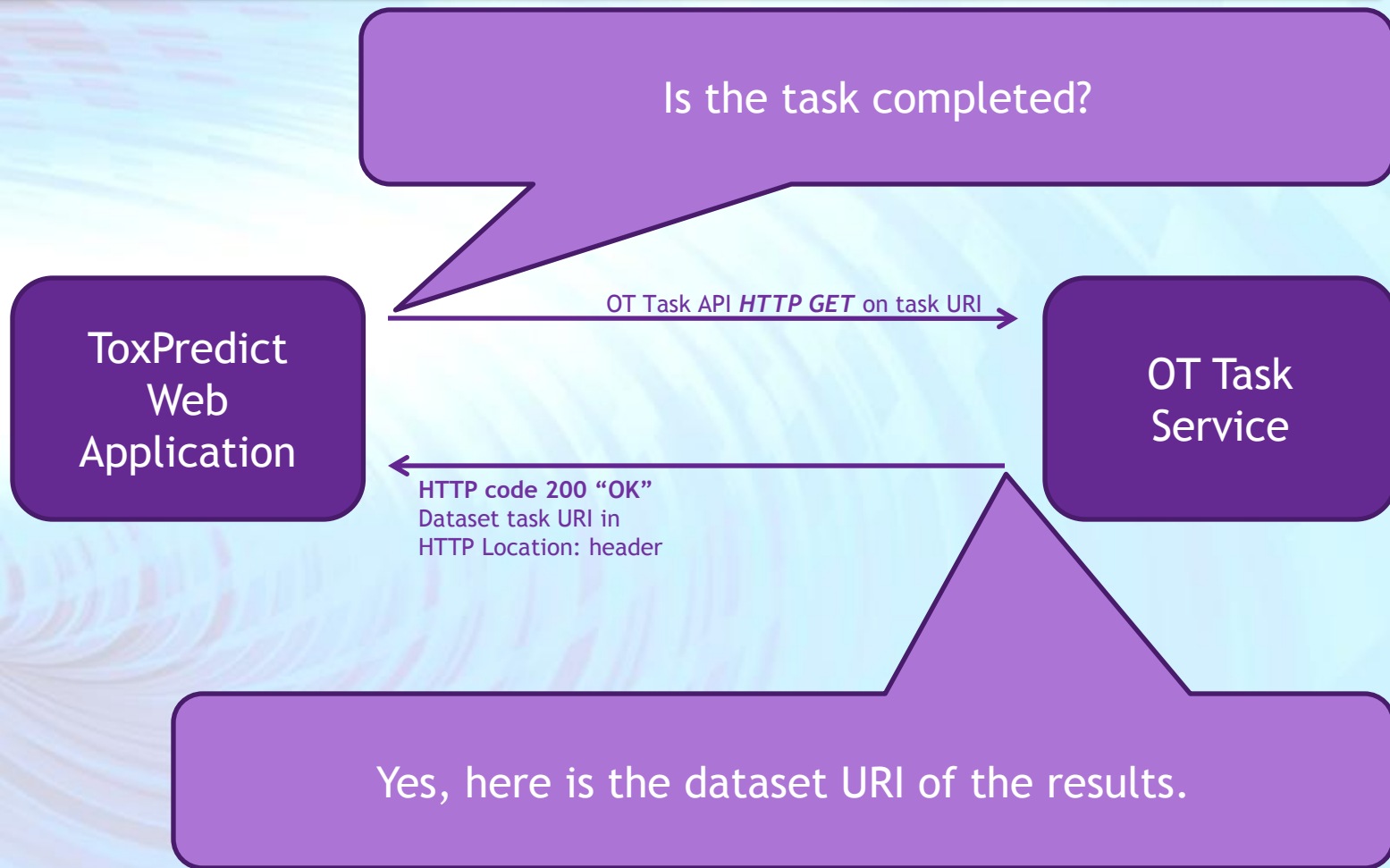
ToxPredict: Step 4 (behind the scenes)



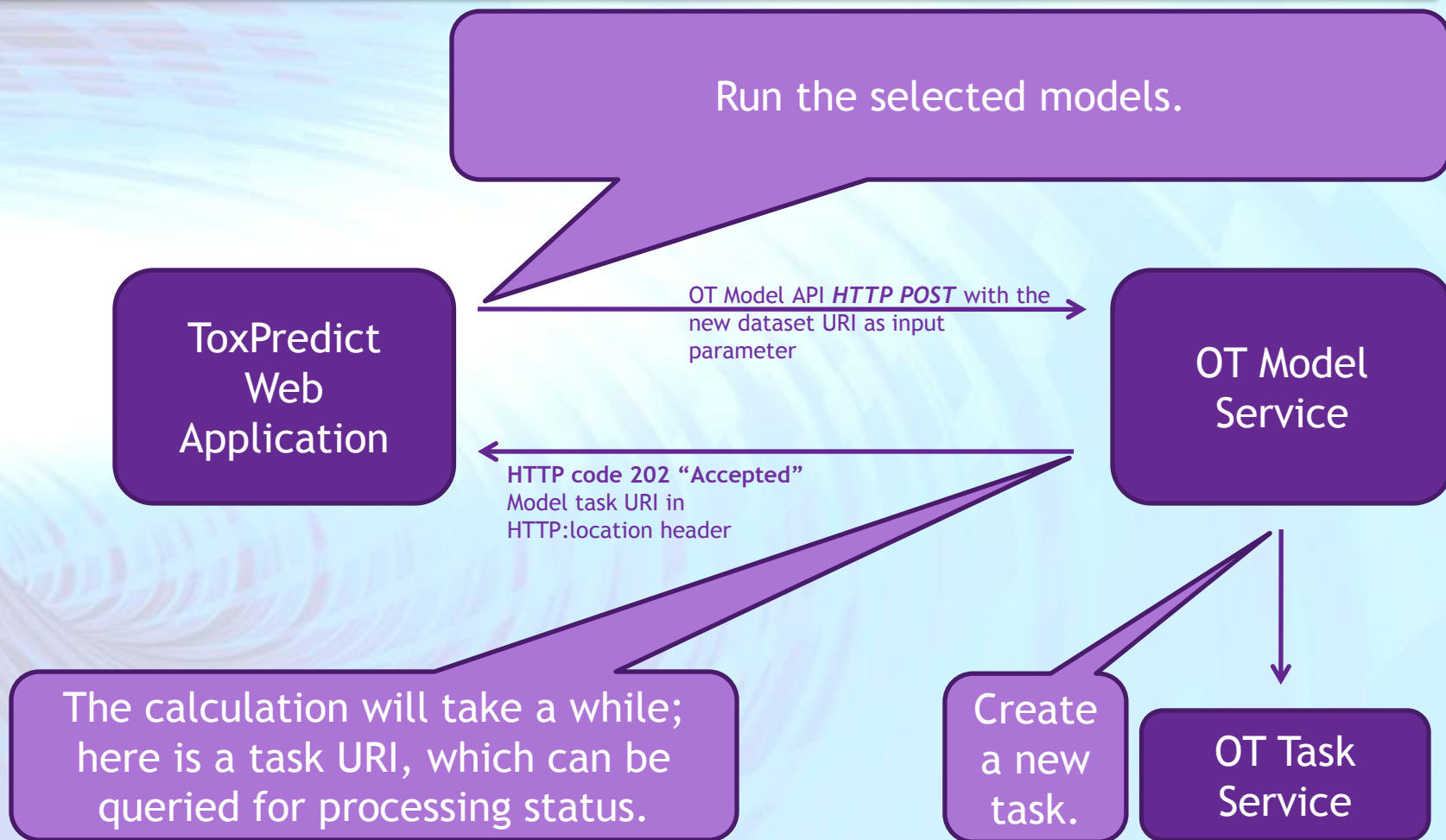
ToxPredict: Step 4 (behind the scenes)



ToxPredict: Step 4 (behind the scenes)

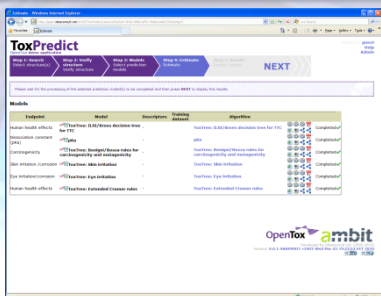


ToxPredict: Step 4 (behind the scenes)



... the remaining part of the Estimation step has been described already in slides 15 and 16

ToxPredict: Step 5 (Display results)



Retrieve calculation results from the final dataset URI, obtained in Step 4 (Estimation).

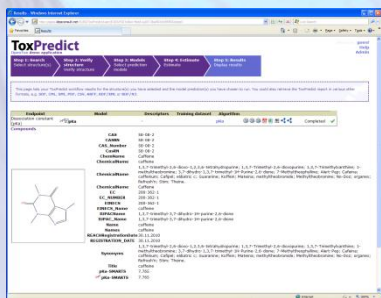


ToxPredict Web Application

OT Dataset Service

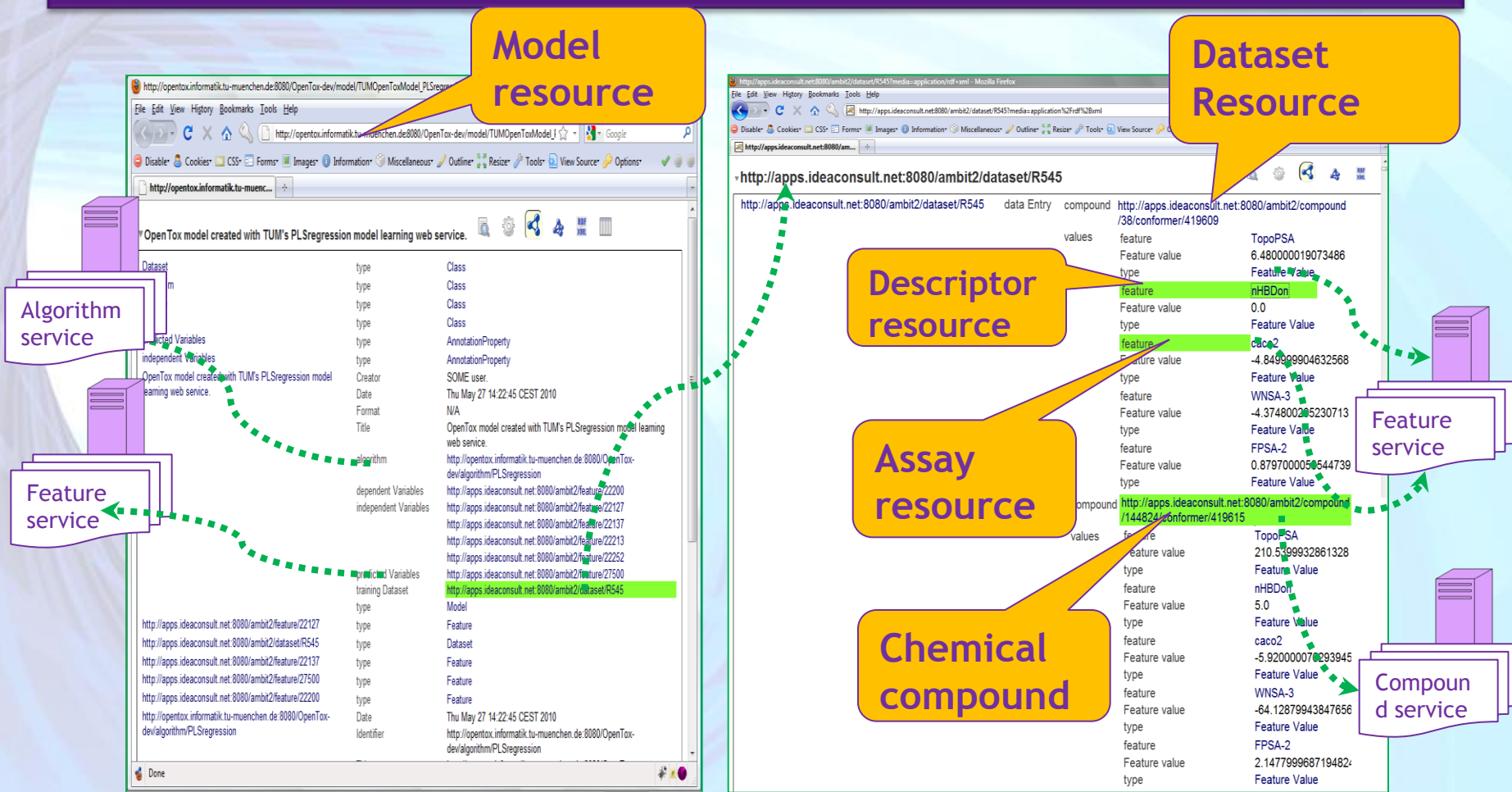
OT Dataset API *HTTP GET*

application/rdf+xml



Here is the dataset content in RDF format, according to OpenTox.owl and containing estimation results, as well as compound identifiers and experimental data.

Linked resources: Compound, Algorithm, Model, Dataset, Features



OpenTox model created with TUM's PLSRegression model learning web service.

Dataset	type	Class
	type	Class
	type	Class
	type	Class
	type	Class
linked Variables	type	AnnotationProperty
independent Variables	type	AnnotationProperty
OpenTox model created with TUM's PLSRegression model learning web service.	Creator	SOME user.
	Date	Thu May 27 14:22:45 CEST 2010
	Format	N/A
	Title	OpenTox model created with TUM's PLSRegression model learning web service.
	algorithm	http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/PLSRegression
	dependent Variables	http://apps.ideaconsult.net:8080/ambit2/feature/22200
	independent Variables	http://apps.ideaconsult.net:8080/ambit2/feature/22127
		http://apps.ideaconsult.net:8080/ambit2/feature/22137
		http://apps.ideaconsult.net:8080/ambit2/feature/22213
		http://apps.ideaconsult.net:8080/ambit2/feature/22252
		http://apps.ideaconsult.net:8080/ambit2/feature/27500
	training Dataset	http://apps.ideaconsult.net:8080/ambit2/dataset/R545
	type	Model
http://apps.ideaconsult.net:8080/ambit2/feature/22127	type	Feature
http://apps.ideaconsult.net:8080/ambit2/dataset/R545	type	Dataset
http://apps.ideaconsult.net:8080/ambit2/feature/22137	type	Feature
http://apps.ideaconsult.net:8080/ambit2/feature/27500	type	Feature
http://apps.ideaconsult.net:8080/ambit2/feature/22200	type	Feature
http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/PLSRegression	Date	Thu May 27 14:22:45 CEST 2010
	Identifier	http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/PLSRegression

<http://apps.ideaconsult.net:8080/ambit2/dataset/R545>

data Entry	compound	http://apps.ideaconsult.net:8080/ambit2/compound/38/conformer/419609
values	feature	TopoPSA
	Feature value	6.48000019073486
	type	Feature Value
	feature	nHBDon
	Feature value	0.0
	type	Feature Value
	feature	caco2
	Feature value	-4.84909904632568
	type	Feature Value
	feature	WNSA-3
	Feature value	-4.374800205230713
	type	Feature Value
	feature	FPSA-2
	Feature value	0.8797000050544739
	type	Feature Value
	compound	http://apps.ideaconsult.net:8080/ambit2/compound/144824/conformer/419615
values	feature	TopoPSA
	Feature value	210.5399932861328
	type	Feature Value
	feature	nHBDOn
	Feature value	5.0
	type	Feature Value
	feature	caco2
	Feature value	-5.920000070293945
	type	Feature Value
	feature	WNSA-3
	Feature value	-64.12879943847656
	type	Feature Value
	feature	FPSA-2
	Feature value	2.147799968719482
	type	Feature Value

Linked resources: Compound, Algorithm, Model, Dataset, Features

Dataset Resource

Descriptor resource

Assay resource

Chemical compound

http://apps.ideaconsult.net:8080/ambit2/dataset/R545

data Entry	compound	http://apps.ideaconsult.net:8080/ambit2/compound/38/conformer/419609
feature	TopoPSA	
Feature value	6.48000019073486	
type	Feature Value	
feature	nHBDon	
Feature value	0.0	
type	Feature Value	
feature	caco2	
Feature value	849999904632568	
type	Feature Value	
feature	WNSA-3	
Feature value	-374800205230713	
type	Feature Value	
feature	FPQA-2	
Feature value	0.8797000050544739	
type	Feature Value	
compound	http://apps.ideaconsult.net:8080/ambit2/compound/144824/conformer/419615	
feature	TopoPSA	
Feature value	210.5399938861328	
type	Feature Value	
feature	nHBDon	
Feature value	5.0	
type	Feature Value	
feature	caco2	
Feature value	-5.920000076293945	
type	Feature Value	
feature	WNSA-3	
Feature value	-64.12879943847856	
type	Feature Value	
feature	FPQA-2	
Feature value	2.147799968719482	
type	Feature Value	

http://apps.ideaconsult.net:8080/ambit2/feature/22213

Name of the algorithm	type	Class
	type	Class
Numeric Feature	type	Class
	subClassOf	
Source	type	ObjectProperty
Units	type	DatatypeProperty
nHBDon	sameAs	http://www.blueobelisk.org/ontologies/cheminformatics-algorithms/#hBondDonors
	Title	nHBDon
	Source	http://apps.ideaconsult.net:8080/ambit2/algorithm/org.openscience.cdk.qsar.descriptors.molecular.HBondDonorCountDescr
	Units	
	type	Numeric Feature

Blue Obelisk algorithms ontology

Regression
Classification
Quantum
Chemistry
Descriptors, etc.

OpenTox algorithm types ontology

Toxicology related ontologies

http://apps.ideaconsult.net:8080/ambit2/feature/22200

Numeric Feature	type	Class
	type	Class
	subClassOf	
Source	type	ObjectProperty
Units	type	DatatypeProperty
caco2	Title	caco2
	Source	c049884m_caco2-training_set.sdf
	Units	
	type	Numeric Feature
	sameAs	Gastrointestinal absorption